Población, muestra y variable estadística

La **estadística** es la parte de las Matemáticas que estudia cómo recopilar y resumir gran cantidad de información para extraer conclusiones.

La **población** de un estudio estadístico es el conjunto de elementos objeto de estudio. Cada elemento se denomina **individuo**. Cuando el número de individuos de la población es muy grande, tomamos una parte de ésta, denominada **muestra.** La muestra es un subconjunto de la población y tiene que ser representativa de la misma.

La **variable estadística** es la propiedad o característica de la población que estamos interesados en estudiar. Puede ser *cualitativa* o *cuantitativa.*

- Las variables **cualitativas** toman valores no numéricos.
- Las variables **cuantitativas** toman valores numéricos. Entre ellas, distinguimos dos tipos: *discretas* y *continuas.*
  - Las variables cuantitativas **discretas** no pueden tomar valores intermedios entre dos valores posibles consecutivos.
  - Las variables cuantitativas **continuas** pueden tomar valores intermedios entre dos valores tan próximos como deseemos.

Estudio estadístico	*Población*	*¿Es necesario tomar muestra?*	*Variable estadística*	*Tipo de variable*
Color del coche de los ciudadanos	Coches de los ciudadanos	Sí	Color	Cualitativa
Altura de los alumnos de la clase	Alumnos de la clase	No	Altura	Cuantitativa continua
Edad de los miembros de una familia	Miembros de la familia	No	Edad	Cuantitativa discreta

Ejercicios:

1.- Indica cuál es la población de cada uno de los siguientes estudios estadísticos y di si es conveniente tomar muestra.

Estudio estadístico	*Población*	*Muestra*
Goles marcados por cada jugador de un equipo		
Comida preferida por los clientes de un restaurante		
Talla de zapato de los miembros de una familia		
Número de hermanos de los habitantes de una ciudad		

2.- Identifica las variable cualitativas y las cuantitativas:

Variable	*Tipo*	
	Cualitativa	*Cuantitativa*
Número de mesas de cada aula		
Longitud de las calles de una ciudad		
Partido más votado en unas elecciones		
Color del pelo de los caballos		

3.-Escribe:

a) Tres ejemplos de variables cualitativas.

b) Tres ejemplos de variables cuantitativas discretas.

c) Tres ejemplos de variables cuantitativas continuas.

4.- Di si las siguientes afirmaciones son verdaderas o falsas. Escribe la frase correcta:

- Para realizar un estudio estadístico se debe investigar a toda la población objeto de estudio.
- La propiedad o característica de la población que queremos estudiar se denomina variable estadística.
- Una muestra es una parte de la población que se desea estudiar.
- Las variables que toman valores no numéricos son variables cualitativas.
- La variable superficie de las viviendas de una ciudad es una variable cuantitativa discreta.
- La variable número de letras de las palabras de un texto es una variable cuantitativa continua.

5.- Completa el cuadro:

Estudio estadístico	*Población*	*¿Se necesita muestra?*	*Variable estadística*	*Tipo de variable*
Proyecciones de una película en los cines de una ciudad				
Distancia del colegio a las casas de los alumnos de una escuela				
				Cualitativa
		No		Cuantitativa discreta
Marca de leche preferida por los ciudadanos europeos.				
				Cuantitativa continua

Frecuencias absoluta, relativa y acumuladas

La **frecuencia absoluta**, f_i *de* un valor x_i de una variable estadística es el número de veces que tomamos dicho valor.

La **frecuencia relativa**, h_i, de un valor x_i determinado de una variable estadística es igual al cociente entre la frecuencia absoluta f_i del valor y el número n de individuos de la población o muestra:

$$f_i = \frac{h_i}{n}$$

$$\% = h_i \cdot 100$$

La **frecuencia absoluta acumulada,** F_i correspondiente a un valor x_i es la suma de las frecuencias absolutas de los valores menores o iguales que el dado:

$$F_i = \sum f_i = f_1 + f_2 + \ldots + f_n$$

La **frecuencia relativa acumulada,** H_i correspondiente a un valor x_i es la suma de las frecuencias relativas de los valores menores o iguales que el dado:

$$H_i = \sum h_i = h_1 + h_2 + \ldots + h_n$$

Ejemplo: *xi: número de hijos*

fi: número de parejas que tienen ese número de hijos

22031	23332	12213	23314
24313	24223	12332	32413
33223	31520	52223	31422
32333	24326	23224	42132
22211	31224	35241	32100
12134	22213		

xi	*fi*	*Fi*	*hi*	*Hi*	*%*
0	4	4	0,036	0,036	36
1	18	22	0,164	0,2	16,4
2	41	63	0,373	0,573	37,3
3	32	95	0,291	0,864	29,1
4	11	106	0,1	0,964	10,0
5	3	109	0,027	0,991	2,7
6	1	110	0,009	1	0,9
	110		1,000		100

Datos agrupados en clases o intervalos:

Cuando en una distribución estadística la variable es continua o el número de valores que toma la variable es muy grande, conviene elaborara una tabla de frecuencias agrupándolas en intervalos. Para ello:

- El número de clases se determina calculando la raíz cuadrada del número de datos y redondeando al entero más próximo.
- La amplitud de cada intervalo se determina:
  - Localizamos los valores extremos *a y b*, y se halla su diferencia: $r = b - a$

- Dividimos el recorrido (r) entre el número de clases. Conviene redondear la amplitud para trabajar con valores cómodos.

Se llama **marca de clase** el valor medio entre los extremos de cada clase.

Ejemplo:

A continuación indicamos las estaturas de 40 adolescentes:

168	160	167	175	175
167	168	158	149	160
178	166	158	163	171
162	165	163	156	174
160	165	154	163	165
161	162	166	163	159
170	165	150	167	164
165	173	164	169	170

- Menor = 149
- Mayor = 178
- R = 178 – 149 =29
- Nº de intervalos: $\sqrt{40}$

  Tomamos 6 intervalos

  Amplitud =29/6

  Redondeamos a cinco intervalos

Intervalos	Marcas de clase	Frecuencias
[148,5-153,5)	151	2
[153,5-158,5)	156	4
[158,5-163,5)	161	11
[163,5-168,5)	166	14
[168,5-173,5)	171	5
[173,5-178,5)	176	4

TABLAS DE FRECUENCIAS.

1. La serie de datos siguiente informa del número de meses que tenían los bebes de un grupo cuando empezaron a andar solos:
   12, 14, 9, 16, 15, 11, 14, 13, 15, 14, 12, 17, 14, 15, 14, 12, 10, 12, 16, 15, 14, 18, 13, 14, 15, 14
   Ordena los datos y agrúpalos en una tabla de frecuencias absolutas, relativas y acumuladas.

2. El tiempo de espera (en minutos) en una parada de guagua de un grupo de personas ha sido:
   2, 15, 7, 9, 4 , 3, 4, 6, 8, 12, 2, 1, 4, 6, 16, 13, 20, 2, 15, 6, 4, 3, 8, 9, 3, 1, 5, 6, 8, 15, 7, 8, 5, 6, 9, 12, 5, 6, 4, 7
   a) Resume los datos en una tabla de datos agrupados.
   b) Calcula las frecuencias.

3. Los jugadores de un equipo de fútbol tienen las siguientes edades:

19	21	25	23	28	23	18	26
23	24	20	27	26	28	19	25
20	22	23	18	27	29	21	26

   a) Resume los datos en una tabla y halla las frecuencias.
   b) ¿Cuántos jugadores tienen menos de 24 años? ¿
   y más de 27 años?

4. Un biólogo, que realiza un estudio sobre la longitud de las musarañas que viven en un bosque, ha encontrado los siguientes datos (en cm):

5,42	6,22	8,42	7,54	6,44	6,76
5,90	6,18	7,16	6,80	7,32	8,12
6,79	7,12	8,21	8,13	7,25	7,34
5,56	8,32	7,45	7,43	6,87	7,10

   Construye una tabla de datos agrupados. Halla la marca de clase y las frecuencias.

GRÁFICOS ESTADÍSTICOS

Las representaciones gráficas deben conseguir que un simple análisis visual ofrezca la mayor información posible. Según el tipo del carácter que estemos estudiando, usaremos una representación gráfica u otra.

A) DIAGRAMAS DE BARRAS

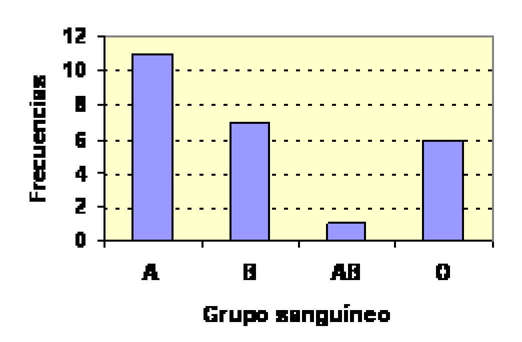
Es un gráfico sobre ejes cartesianos en el que distribuimos en el eje X o eje de abscisa:

- Las modalidades si el carácter es cualitativo
- Los valores si la variable es no agrupada

Sobre ellos se levantan barras o rectángulos de igual base (que no se solapen) cuya altura sea proporcional a sus frecuencias. También pueden representarse horizontalmente, intercambiando los ejes.

Ejemplo 1. Un estudio hecho en un conjunto de 25 varones con objeto de determinar su grupo sanguíneo ha conducido a los siguientes resultados:

Modalidad	Frecuencia absoluta
A	11
B	7
O	6
AB	1
	25



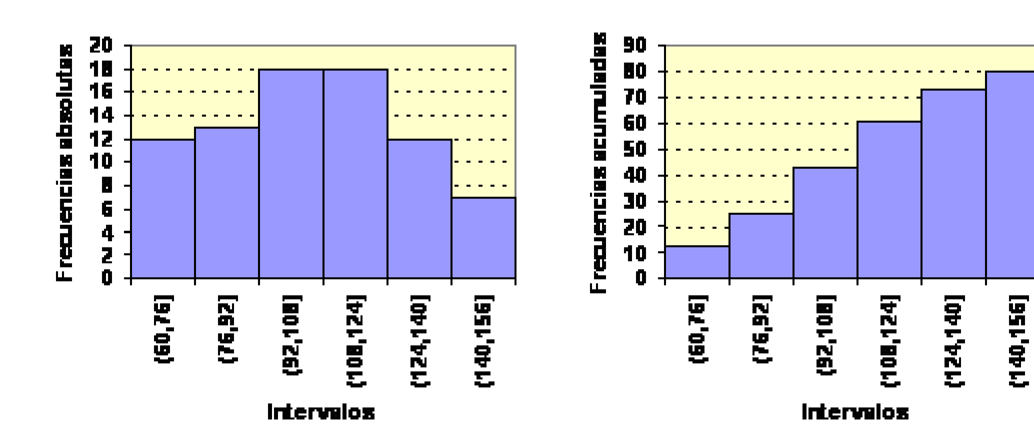
B) HISTOGRAMAS

Se utiliza con variables agrupadas en intervalos, representando en el eje X los intervalos de clase y levantando rectángulos contiguos de base la longitud de los distintos intervalos y de altura tal que el área sea proporcional a las frecuencias representadas.

Ejemplo:

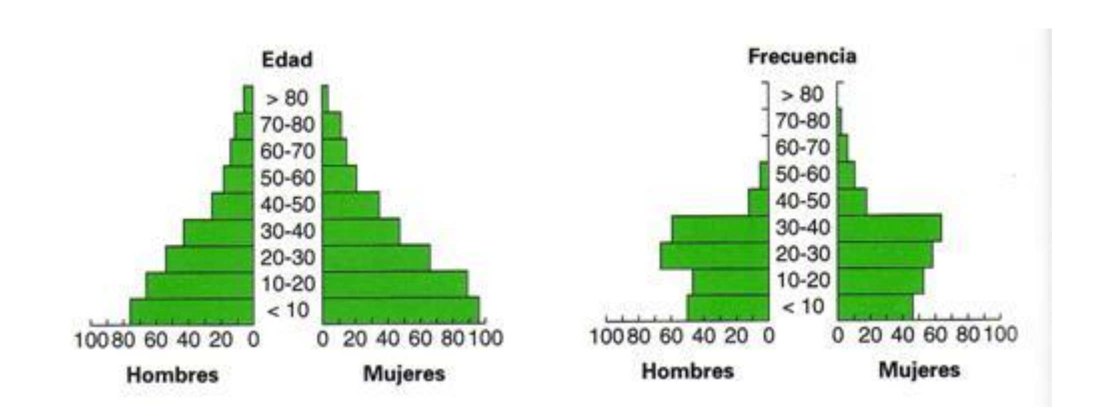
El número de personas que viven en cada uno de los portales de una gran barriada es:

Intervalos de clase	Marca de clase	Frecuencia absoluta	Frecuencia relativa	Porcentajes	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
(60,76]	68	12	12/80	15%	12	12/80
(76,92]	84	13	13/80	16’25%	25	25/80
(92,108]	100	18	18/80	22’5%	43	43/80
(108,124]	116	18	18/80	22’5%	61	61/80
(124,140]	132	12	12/80	15%	73	73/80
(140,156]	148	7	7/80	8’75%	80	1
		80	1	100%		



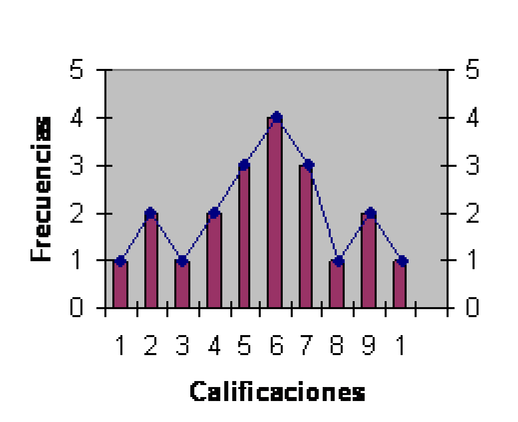
En este caso, todos los intervalos son de la misma longitud, por lo que la altura de cada rectángulo coincide con la frecuencia.

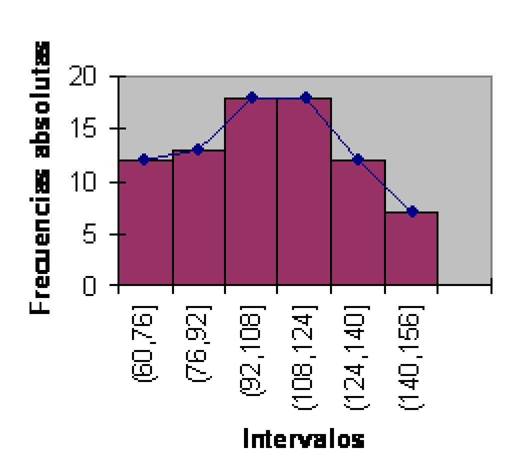
Cuando se realizan representaciones correspondientes a edades de población, cambiamos el eje Y por el eje X para obtener las llamadas **pirámides de población**, que no son más que 2 histogramas a izquierda y derecha, para hombres y mujeres. Veamos un ejemplo:



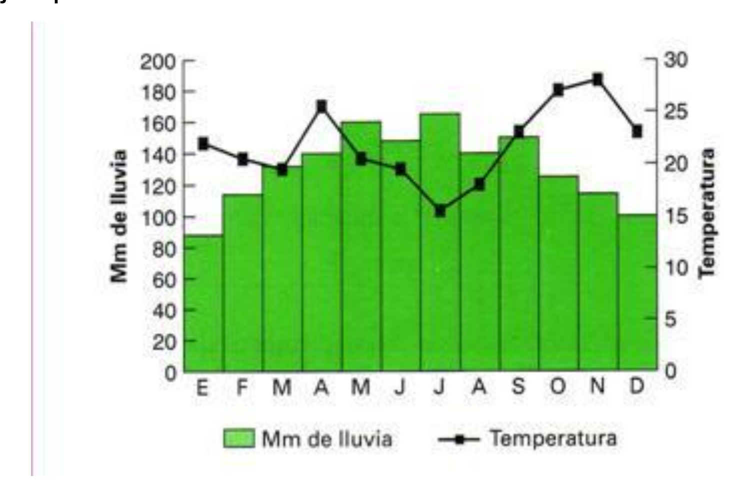
C) POLÍGONOS DE FRECUENCIAS

Son gráficos lineales que se utilizan en el caso de una variable cuantitativa. Para realizar estos polígonos unimos los puntos medios de las bases superiores del diagrama de barras o del histograma según la variable sea agrupada o no agrupada.





Un caso particular de aplicación de los histogramas y los polígonos de frecuencias es el **climograma**, que representa la marcha anual de las temperaturas y de las lluvias medias, sobre un mismo sistema de coordenadas. Veamos un ejemplo:



En el caso de representar las frecuencias acumuladas se unen los puntos medios de las bases superiores del diagrama de barras, si la variable es **no agrupada**, y los vértices superiores derechos de los rectángulos si se trata de una variable **agrupada**.

D) DIAGRAMA DE SECTORES

Son gráficos en los que a cada valor o modalidad se reasigna un sector circular de área proporcional a la frecuencia que representan. Se utilizan si el carácter es cualitativo o cuantitativo discreto no agrupado.

Realicemos el diagrama de sectores del ejemplo 1.



La amplitud de cada sector se obtiene multiplicando la frecuencia relativa del valor de la variable por 360.

E) PICTOGRAMAS

Son gráficos con dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan; dicha frecuencia se suele representar.

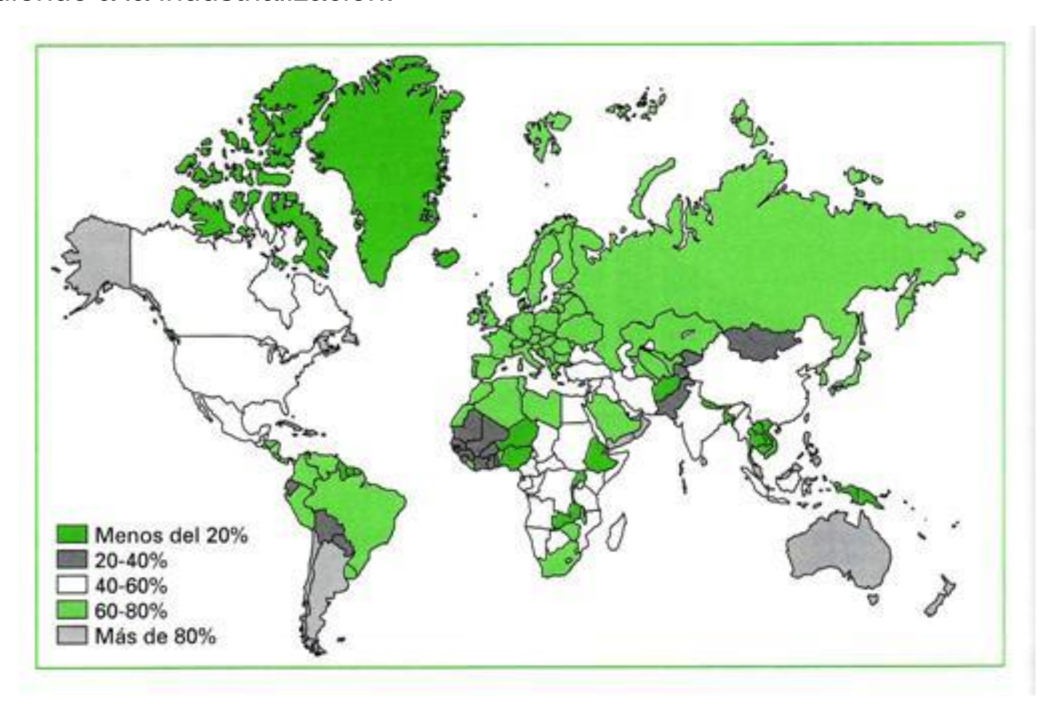
En el siguiente ejemplo hemos representado el número de partidos ganados, perdidos o empatados de un equipo.



F) CARTOGRAMAS

Son gráficos realizados sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.

En el siguiente cartograma observamos la urbanización en el mundo atendiendo a la industrialización.



PARÁMETROS DE CENTRALIZACIÓN.

Los parámetros de centralización o medidas de posición central son números que nos indican alrededor de qué valor se distribuyen los valores de la variable estadística observada.

Media

Es la medida de posición central más utilizada. Para calcularla se utiliza la siguiente expresión:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{n}$$

Veamos cómo se calcula la media, utilizaremos el ejemplo visto en las tablas de frecuencias:

xi	*fi*	*Fi*	*hi*	*Hi*	$x_i \cdot f_i$
0	4	4	0,036	0,036	0
1	18	22	0,164	0,2	18
2	41	63	0,373	0,573	82
3	32	95	0,291	0,864	96
4	11	106	0,1	0,964	44
5	3	109	0,027	0,991	15
6	1	110	0,009	1	6
	110		1,000		261

$$\bar{x} = \frac{261}{110} = 2,37$$

Cuando los datos están agrupados en clases o intervalos tomamos la marca de clase.

Intervalos	Marcas de clase	Frecuencias	$c_i \cdot f_i$
[148,5-153,5)	151	2	302
[153,5-158,5)	156	4	624
[158,5-163,5)	161	11	1771
[163,5-168,5)	166	14	2324
[168,5-173,5)	171	5	855
[173,5-178,5)	176	4	704
		40	6580

$$\bar{x} = \frac{6580}{40} = 164{,}5$$

Mediana

La mediana es el dato que ocupa la posición intermedia de la distribución, está después del 50% de los datos y precediendo al otro 50%.

Ejemplos:

1.- Supongamos que un alumno ha obtenido las siguientes notas en los exámenes de Matemáticas que ha realizado en el curso:

5, 6, 4, 7, 8, 8, 9, 7, 9, 5, 7

Ordenamos estos valores de menor a mayor y observamos el valor que ocupa la posición central:

4, 5, 5, 6, 7, **7**, 7, 8, 8, 9, 9

La mediana es *Me* = 7

Si el número de datos es par, se toma como mediana la media aritmética de los datos que ocupan la posición central.

2.- Ahora queremos hallar la mediana de las notas obtenidas por los alumnos de un grupo en un examen. Las notas están agrupadas:

x_i	f_i	F_i
4	5	5
5	6	11
6	8	19
8	4	23
9	2	25
	25	

Cómo hay 25 datos, el dato central es el que ocupa el lugar 13, que pertenece al valor cuya frecuencia acumulada es mayor que 13, es decir la tercera fila de datos de la tabla. Por tanto, *Me* =6

3.-

Intervalos	Marcas de clase	f_i	F_i
[148,5-153,5)	151	2	2
[153,5-158,5)	156	4	6
[158,5-163,5)	161	11	17
[163,5-168,5)	166	14	31
[168,5-173,5)	171	5	36
[173,5-178,5)	176	4	40
		40	

Hay 40 datos, los datos centrales están en la posición 20 y 21, que pertenece a la 4ª fila de datos. La clase correspondiente a esta fila se llama clase mediana. En este caso [163,5 – 168,5).

Moda.

La moda es el valor de la variable que tiene más frecuencia, es decir, que se ha obtenido más veces.

En los ejemplos anteriores:

1.- *Mo* = 7

2.- *Mo* = 6

3.- La moda es la marca de clase del intervalo de mayor frecuencia.

Mo = 166

La moda se utiliza cuando no conviene o no se puede calcular ni la media y ni la mediana. Podemos tener distribuciones unimodales, bimodales,…

Parámetros de dispersión.

Los parámetros de dispersión son medidas que indican hasta qué punto la variable estadística toma valores próximos o alejados de las medidas de posición central.

Recorrido o rango:

Es la diferencia entre el valor mayor y el valor menor de la variable.

Desviaciones respecto a la media.

Se llama desviación respecto a la media de un dato x_i a la diferencia $x_i - \bar{x}$.

Varianza:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{n}$$

Desviación típica:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2 \cdot f_i}{n}}$$

Ejemplo:

1.

x_i	f_i	F_i	$x_i \cdot f_i$	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 \cdot f_i$
4	5	5	20	-1,92	3,6864	18,432
5	6	11	30	-0,92	0,8464	5,0784
6	8	19	48	0,08	0,0064	0,0512
8	4	23	32	2,08	4,3264	17,3056
9	2	25	18	3,08	9,4864	18,9728
	25		148			59,84

$$\overline{x} = \frac{148}{25} = 5{,}92$$

$$s^2 = \frac{59{,}84}{25} = 2{,}3936$$
$$s = \sqrt{2{,}3936} = 1{,}547$$

Ejercicios:

De los ejercicios 1, 2, 3 y 4. Calcular la media, mediana, moda y desviación típica.

5. En una residencia, se ha tomado la muestra siguiente de las edades de los residentes:

   76, 82, 85, 81, 79, 82, 84, 90, 87, 91, 86, 83, 92, 85, 81, 83, 75, 77 y 79

   a) Halla la media, mediana y moda de la distribución.
   b) Calcula la desviación típica.

6. Se ha pasado una encuesta a 60 estudiantes de 4º ESO para investigar el gasto semanal en actividades de ocio. Los resultados han sido:

gasto	Nº de estudiantes
[0,6)	4
[6,12)	12
[12, 18)	25
[18, 24)	10
[24, 30)	5

   a) Completa la tabla de frecuencias.
   b) Representa los datos en un histograma y en un diagrama de sectores.
   c) ¿Cuántos alumnos gastan menos de 12 €?
   d) ¿Qué porcentaje de alumnos gasta más de 18 €?
   e) Calcula la media, mediana y moda.
   f) Calcula la desviación típica.

7. Los salarios mensuales de cinco empleados de una empresa son 900€, 1000€, 1500€, 2000€ y 2100€. Se incorpora a la empresa un nuevo empleado con 3000€ de salario mensual.

   a) Halla la media de los salarios de los cinco empleados iniciales.
   b) Calcula la media de los salarios después de la incorporación del nuevo empleado.
   c) ¿Te parece que la media es una buena medida de posición central en los casos anteriores?

d) ¿Qué medida de centralización te parece más adecuada?. Calcúlala en los dos casos.

8. A un grupo de 30 personas se les ha tomado el número de pulsaciones por minuto (ritmo cardiaco) obteniéndose los siguientes resultados:

87	85	61	51	64	75	80	70	69	82
80	79	82	74	90	76	72	73	63	65
67	71	88	76	68	73	70	76	71	86

a) Agrupa los datos en cinco intervalos y construye la tabla de frecuencias.

b) Representa gráficamente esta distribución.

c) Calcula la media, mediana y moda.

d) Calcula la desviación típica.